**TWITTER DATA &**

**SENTIMENT ANALYSIS**

**Name: Anshika Mittal**

**University Roll No: 2012505**

Problem Statement:

Collection of twitter data and analysing the sentiments expressed in the tweet whether it is positive, negative or neutral.

Introduction:

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users [24] - out of which 100 million are active users and half of them log on twitter on a daily basis. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream

The objective of the project is to stream the live tweets from the twitter account, accessing all the information related to the tweets Analysing the tweets and the data the we have collected and finally analysing the sentiments of the tweet classifying it into three categories: positive, negative or neutral.

Streaming Twitter Data:

We have started with making the developer account for twitter and created an application Analyse sentiment tweet to use its credentials for accessing the tweets. We have made use of tweepy package in python to stream the live tweets and also for the authentication of the tweets that we are going to stream. We have made use of Stream listener to listen all the tweets. OAuthHandler is used for the authentication of the tweets. We are Fetching the tweets on the basis of the topic list. Topic list is the list of the keyword for which we want to stream the tweets. All the tweets containing that keyword will be streamed.

Accessing & publishing tweets:

We have used cursor and pagination technique over here. Pagination is the way of numbering of tweets that we have accessed. Using the tweepy package we can access the tweets which our on the user’s timeline, we can access the tweets on its timeline which are posted by the people he/she follows and we can also access the friend list of the user. While calling the function we have to call the function by pass the number of tweets we want to see. After the function is called we will be able to see all the tweets on the console for the user we have mentioned in the topic list. We have also made use of filter function to filter the tweet since some might be useful and some might be not. Here we are finished with the streaming of the tweets. I have used on\_error method which check if we are doing something unauthenticated. It alerts us if we do something which is unauthorized.

Analysing Twitter Data:

After retrieving the data we will analyse the twitter data. We will use pandas and numpy for analysis. The class twitter client will have the function which will return the twitter client. Here we will retrieve the additional data which is related to the tweets. We will use the screen name which is the handle name in the twitter to retrieve the information. We will create a data frame to organize the tweets and other information. We are saving the id in the data frame by creating an array for id. Also with id we can add length of the tweet, date of the tweet, like od the tweet, retweets and source of the tweet to the data frame.

Visualising Twitter Data:

All the data that has been retrieved using data frame can be visualized. Now that we have stored everything in the data frame it will be easy to calculate the mean and the average. We can calculate the most liked tweet, what is the average of the likes on the number of tweets. We can perform endless operations using this data. We have also used time series to plot the graph for likes and retweets to analyse what does user like and what does not.

Sentiment Analysis:

Here text blob library has been used in which sentimental analyser is already trained. We have made use of regular expression to remove the extra characters or hyper links ,since they do not add something to the sentiment expressed in the tweet. We will check the polarity of the tweet for analysis of the sentiment. If the polarity of the tweet is greater than 0 then the tweet is **positive** ,if the polarity is equal to 0 then the tweet is **neutral** and if the polarity of the tweets is less than 0 then the tweets is **negative.** This is how we can collect the data from Twitter and can analyse the sentiment expressed in the tweet.